86e The Human Microbiome

Jeffrey I. Gordon, Rob Knight

The technologies that allowed us to decipher the human genome have revolutionized our ability to delineate the composition and functions of the microbial communities that colonize our bodies and make up our *microbiota*. Each body habitat, including the skin, nose, mouth, airways, gastrointestinal tract, and vagina, harbors a distinctive community of microbes. Efforts to understand our microbiota and its collection of microbial genes (our *microbiome*) are changing our views of "self" and deepening our understanding of many normal physiologic, metabolic, and immunologic features and their interpersonal and intrapersonal variations. In addition, this area of research is beginning to provide new insights into diseases not previously known to have microbial "contributors" and is suggesting new strategies for treatment and prevention. Key terms used in the discussion of the human microbiome are defined in Table 86e-1.

We are holobionts—collections of human and microbial cells that function together in an elaborate symbiosis. The aggregate number of microbial cells in our microbiota exceeds the number of human cells in our adult bodies by up to 10-fold, and each healthy adult is estimated to harbor 10^5–10^6 microbial genes, in contrast to ~20,000 *Homo sapiens* genes. Members of our microbiota can function as mutualists (i.e., both host and microbe benefit from each other's presence), as commensals (one partner benefits; the other is seemingly unaffected), and as potential or overt pathogens (one partner benefits; the other is harmed). Many clinicians view pathogens as individual microbial species or strains that can elicit disease in susceptible hosts. An emerging, more ecologic view is that pathogens do not function in isolation; rather, their invasion, emergence, and effects on the host reflect interactions with other members of a microbiota. An even more expansive view is that multiple organisms in a community conspire to

TABLE 86e-1 GLOSSARY OF TERMS USED IN DISCUSSION OF THE HUMAN MICROBIOME

Term	Definition
Culture-independent analysis	A type of analysis in which the culture of microbes is not required but rather information is extracted directly from environmental samples
Diversity (alpha and beta)	*Alpha* diversity measures the effective number of species (kinds of organisms) at the level of individual habitats, sites, or samples. *Beta* diversity measures differences in the number of kinds of organisms across habitats, sites, or samples.
Domains of life	The three major branches of life on Earth: the Eukarya (including humans), the Bacteria, and the Archaea
Dysbiosis	Any deleterious condition arising from a structural and/or functional aberration in one or more of the host organism's microbial communities
Gnotobiotics	The rearing of animals under sterile (germ-free) conditions. These animals can subsequently be colonized at various stages of the life cycle with defined collections of microbes.
Holobiont	The biologic entity consisting of a host and all its internal and external symbionts, their gene repertoires, and their functions
Human microbiome	In ecology, *biome* refers to a habitat and the organisms in it. In this sense, the human *microbiome* would be defined as the collection of microorganisms associated with the human body. However, the term *microbiome* is also used to refer to the collective genomes and genes present in members of a given microbiota (see "Microbiota," below), and the human *metagenome* is the sum of the human genome and microbial genes (microbiome). A *core human microbiome* is defined as everything shared in a given body habitat among all or the vast majority of human microbiomes. A core microbiome may include a common set of genomes and genes encoding various protein families and/or metabolic capabilities. Microbial genes that are variably represented in different humans may contribute to distinctive physiologic/metabolic phenotypes.
Metagenomics	An emerging field encompassing culture-independent studies of the structures and functions of microbial communities as well as the interactions of these communities with the habitats they occupy. Metagenomics includes (1) shotgun sequencing of microbial DNA isolated directly from a given environment and (2) high-throughput screening of expression libraries constructed from cloned community DNA to identify specific functions such as antibiotic resistance (*functional metagenomics*). DNA-level analyses provide the foundation for profiling of mRNAs and proteins produced by a microbiome (*metatranscriptomics* and *metaproteomics*) and for identification of a community's metabolic network (*metametabolomics*).
Microbial source tracking	A collection of methods for assessing the environments of origin for microbes. One method, SourceTracker, uses a Bayesian approach to identify each bacterial taxon's origins and estimates the proportions of each community made up by bacteria originating from different environments.
Microbiota	A microbial community—including Bacteria, Archaea, Eukarya, and viruses—that occupies a given habitat
Pan-genome	The group of genes found in genomes that make up a given microbial phylotype, including both *core* genes found in all genomes and variably represented genes found in a subset of genomes within the phylotype
Phylogenetic analysis	Characterization of the evolutionary relationships between organisms and their gene products
Phylogenetic tree	A "tree" in which organisms are shown according to their relationships to hypothetical common ancestors. When built from molecular sequences, the branch lengths are proportional to the amount of evolutionary change separating each ancestor–descendant pair.
Phylotype	A phylogenetic group of microbes, currently defined by a threshold percentage identity shared among their small-subunit rRNA genes (e.g., ≥97% for a species-level phylotype)
Principal coordinates analysis	An ordination method for visualizing multivariate data based on the similarity/dissimilarity of the measured entities (e.g., visualization of bacterial communities based on their UniFrac distances; see "UniFrac," below)
Random Forests analysis/ machine learning	*Machine learning* is a collection of approaches that allow a computer to learn without being explicitly programmed. *Random Forests* is a machine-learning method for classification and regression that uses multiple decision trees during a training step.
Rarefaction	A procedure in which subsampling is used to assess whether all the diversity present in a given sample or set of samples has been observed at a given sampling depth and to extrapolate how much additional sampling would be needed to observe all the diversity
Resilience	A community's ability to return to its initial state after a perturbation
Shotgun sequencing	A method for sequencing large DNA regions or collections of regions by fragmenting DNA and sequencing the resulting smaller sections
Succession (primary and secondary)	*Succession* (in an ecologic context) refers to changes in the structure of a community through time. *Primary* succession describes the sequence of colonizations and extinctions that occur in a new habitat. *Secondary* succession refers to changes in community structure after a disturbance.
UniFrac	A measure of the phylogenetic dissimilarity between two communities, calculated as the unshared proportion of the phylogenetic tree containing all the organisms present in either community